

## Genetic Variation and Evolutionary Origin of the Direct Repeat Locus of *Mycobacterium tuberculosis* Complex Bacteria

J. D. A. VAN EMBDEN,<sup>1\*</sup> T. VAN GORKOM,<sup>1</sup> K. KREMER,<sup>2</sup> R. JANSEN,<sup>3</sup>  
B. A. M. VAN DER ZEIJST,<sup>3</sup> AND L. M. SCHOULS<sup>1</sup>

Department of Bacteriology of the Research Laboratory for Infectious Disease,<sup>1</sup> and Diagnostic Laboratory for Infectious Diseases and Perinatal Screening,<sup>2</sup> National Institute of Public Health and the Environment, 3720 BA Bilthoven, and Veterinary Faculty, Infectious Diseases and Immunology, Department of Bacteriology, 3508 TD Utrecht,<sup>3</sup> The Netherlands

Received 12 November 1999/Accepted 26 January 2000

The direct repeat region in *Mycobacterium tuberculosis* complex strains is composed of multiple direct variant repeats (DVRs), each of which is composed of a 36-bp direct repeat (DR) plus a nonrepetitive spacer sequence of similar size. It has been shown previously that clinical isolates show extensive polymorphism in the DR region by the variable presence of DVRs, and this polymorphism has been used in the epidemiology of tuberculosis. In an attempt to better understand the evolutionary scenario leading to polymorphic DR loci and to improve strain differentiation by spoligotyping, we characterized and compared the DNA sequences of the complete DR region and its flanking DNA of *M. tuberculosis* complex strains. We identified 94 different spacer sequences among 26 *M. tuberculosis* complex strains. No sequence homology was found between any of these spacers and *M. tuberculosis* DNA outside of the DR region or with any other known bacterial sequence. Although strains differed extensively in the presence or absence of DVRs, the order of the spacers in the DR locus was found to be well conserved. The data strongly suggest that the polymorphism in clinical isolates is the result of successive deletions of single discrete DVRs or of multiple contiguous DVRs from a primordial DR region containing many more DVRs than seen in present day isolates and that virtually no scrambling of DVRs took place during evolution. Because the majority of the novel spacer sequences identified in this study were confined to isolates of the rare *Mycobacterium canettii* taxon, the use of the novel spacers in spoligotyping led only to a slight improvement of strain differentiation by spoligotyping.

Bacterial isolates belonging to the *Mycobacterium tuberculosis* complex group of bacteria show an unusually high degree of conservation in housekeeping genes (19, 31). Because the mutation frequency of *M. tuberculosis* is similar to that of other bacteria, these observations have led to the speculation that the *M. tuberculosis* complex isolates presently seen have diverged from a common ancestor not more than 10,000 to 15,000 years ago (19). Much more DNA polymorphism in *M. tuberculosis* complex bacteria has been found to be associated with repetitive DNA, such as transposable elements, and short perfect or imperfect repeats (29, 32). Furthermore, the establishment of the genome sequence of *M. tuberculosis* (5) has led to the disclosure of variation in the presence of multiple regions carrying a variety of different genes (12).

The direct repeat (DR) locus is presently the only well-studied single locus in the genome of *M. tuberculosis* showing considerable strain-to-strain polymorphism (10, 14). This locus is composed of multiple 36-bp DR copies, which are interspersed by nonrepetitive short sequences of about equal length, the so-called spacers (16, 18). Clinical isolates of *M. tuberculosis* and *Mycobacterium bovis* generally differ in the presence or absence of one or more spacers and adjacent DRs. This polymorphism has been exploited to distinguish *M. tuberculosis* complex strains for epidemiological studies and to distinguish the taxons within the group of the *M. tuberculosis* complex, these being *M. tuberculosis*, *M. bovis*, *M. microti*, and *M. canettii*

(9, 13, 15, 18, 22, 34, 37, 39, 40). The function of the DR locus in *M. tuberculosis* is presently unknown. The apparent omnipresence of this unusual region in *M. tuberculosis* complex and the strong sequence conservation of the DRs and spacers among clinical isolates may suggest a biological function of the DR region for the host and that, due to a selective advantage, this region has been maintained in the population. Although no significant homology has been reported between the *M. tuberculosis* DR sequence and DNA of other bacterial genera, loci with similar motifs composed of short repetitive and non-repetitive sequences in other bacterial genera have been found (1, 4, 17, 20, 21, 24–26, 33). In *Haloferax* spp. the locus may be involved in replicon partitioning (26), but in other organisms the function has not been investigated.

Until now, DNA polymorphism in the DR locus has been investigated only in *M. tuberculosis* complex bacteria (10, 14) and in *Streptococcus pyogenes* (17). Isolates of *M. tuberculosis* complex differ in the presence or absence of one or more discrete DNA segments, each consisting of a DR plus the adjacent spacer, the so-called direct variant repeat (DVR). This suggests that homologous recombination between neighboring or distant DRs may lead to deletion of one or more discrete DVRs (10, 14). A similar polymorphism among clinical isolates was observed in the DR locus of *S. pyogenes* (17). Furthermore, the DR region in *M. tuberculosis* has been identified as a hot spot of integration of the insertion element IS6110 (10, 16), and IS6110-associated polymorphism in the DR region has been disclosed, also among outbreak strains (3, 10, 14, 23). To elucidate the mechanism of genetic variation in the DR locus and the evolutionary pathway of the DR locus in *M. tuberculosis* complex bacteria, we compared in this study the sequence of the DR locus and its flanking DNA from a variety of dif-

\* Corresponding author. Mailing address: Department of Bacteriology of the Research Laboratory for Infectious Disease, National Institute of Public Health and the Environment, P.O. Box 1, 3720 BA Bilthoven, The Netherlands. Phone: 31-30-2742113. Fax: 31-30-2744449. E-mail: JDA.van.Embden@rivm.nl.

TABLE 1. Mycobacterial strains used for DNA sequencing

Strain no. <sup>a</sup>	Original designation	Species	Origin	Accession no. <sup>c</sup>	Source or reference
1	NLA009400768	<i>M. tuberculosis</i>	Human	AF189746/AF189747	This laboratory
2	NLA009400982	<i>M. tuberculosis</i>	Human	AF189748/AF189749	This laboratory
3	NLA000016314	<i>M. tuberculosis</i>	Human (previously strain 103)	AF189750/AF189751	Kremer et al. (22)
4	NLA009400319	<i>M. tuberculosis</i>	Human	AF189752/AF189753	This laboratory
5	NLA000016923	<i>M. tuberculosis</i>	Human	AF189754/AF189755	This laboratory
6	NLA009700584	<i>M. tuberculosis</i>	Human	AF189756/AF189757	This laboratory
7	NLA009401014	<i>M. tuberculosis</i>	Human (previously strain 56)	AF189758/AF189759	Kremer et al. (22)
8	H37Rv	<i>M. tuberculosis</i>	Laboratory strain, genome sequenced	Z81331	Cole et al. (5)
9	DR strain 86	<i>M. tuberculosis</i>	Human	Y14045NID	Fang et al. (10)
10	DR strain 257	<i>M. tuberculosis</i>	Human	Y14047NID	Fang et al. (10)
11	DR strain 149	<i>M. tuberculosis</i>	Human	Y14046NID	Fang et al. (10)
12	DR strain 191	<i>M. tuberculosis</i>	Human	Y14048NID	Fang et al. (10)
13	DR strain 93	<i>M. tuberculosis</i>	Human	Y14049NID	Fang et al. (10)
14	CSU#93	<i>M. tuberculosis</i>	Human, genome being sequenced		B. B. Plikaytis
15	NLA009400303	<i>M. tuberculosis</i>	Human	AF189760	This laboratory
16	NLA009400116	<i>M. tuberculosis</i>	Human	AF189818/AF189761	This laboratory
17	NLA000016319	<i>M. tuberculosis</i>	Human (previously strain 111)	AF189762/AF189763	Kremer et al. (22)
18	TN4937	<i>M. tuberculosis</i>	Human	AF189819/AF189820	This laboratory
19	NLA009402008	<i>M. tuberculosis</i>	Human	AF189821/AF189822	Kremer et al. (22)
20	NLA0000000P3	<i>M. bovis</i> BCG	Vaccine strain, The Netherlands	X57835NID	This laboratory
21	NLA0000Russia	<i>M. bovis</i> BCG	Vaccine strain, Russia	AF189823/AF189824	This laboratory
22	TMC401	<i>M. bovis</i>	Bovine	U47864NID	Beggs et al. (2)
23	NLA009401854	<i>M. bovis</i>	Bovine	AF189825/AF189826	D. Brittain
24	NLA009502227	<i>M. tuberculosis</i> complex <sup>b</sup>	Human	AF189827	This laboratory
25	NLA000016240	<i>M. microti</i>	Vole	AF189828/AF189829	Van Soolingen et al. (36)
26	NLA000017727	<i>M. canettii</i>	Human (previously strain SO93)	AF190853	Van Soolingen et al. (37)
27	NLA000013667	<i>M. tuberculosis</i>	Human, patient 1		This laboratory
28	NLA000013668	<i>M. tuberculosis</i>	Human, patient 1		This laboratory
29	NLA009602019	<i>M. tuberculosis</i>	Human, patient 2		This laboratory
30	NLA009602017	<i>M. tuberculosis</i>	Human, patient 2		This laboratory
31	NLA000017877	<i>M. tuberculosis</i>	Human		This laboratory
32	NLA009401835	<i>M. tuberculosis</i>	Human		This laboratory
33	Strain A	<i>M. bovis</i>	Human		Blazquez et al. (3)
34	Strain B	<i>M. bovis</i>	Human		Blazquez et al. (3)

<sup>a</sup> The complete DR regions of strains 1 to 26 were determined. Strains 27 to 34 comprise presumed isogenic pairs, and only part of the DR region was determined.

<sup>b</sup> Phenotype between *M. tuberculosis* and *M. bovis*.

<sup>c</sup> The sequences under two accession numbers were used to obtain the complete sequence of the DR locus.

ferent strains, including pairs of isogenic strains with different spoligotypes. A second objective of this study was to improve the degree of strain differentiation based on DNA polymorphism in the DR region of *M. tuberculosis* by spoligotyping. In this method the whole DR region is amplified and labeled by PCR using DR-specific primers and the presence of any of a set of 43 different spacers is determined by hybridization of the amplified DNA to 43 spacer oligonucleotides, which are covalently linked to a membrane (18). Spoligotypes are expressed as the presence or absence of any of these 43 spacers. The method allows a high throughput, and no cultured cells are needed to differentiate strains (9, 30, 36, 40). A drawback of the presently used method of spoligotyping is the limited discriminative power, compared to other methods such as IS6110 fingerprinting (8, 13, 18, 22, 38, 39). The sequences disclosed in this study to characterize the nature of genetic variation in the DR locus allowed us to evaluate the novel spacers for their potential for improved strain differentiation.

#### MATERIALS AND METHODS

**Bacterial strains.** Unless otherwise stated, all bacterial strains were clinical isolates of *M. tuberculosis* complex. Table 1 presents the strains for which the sequence of the DR region or part thereof was established. For spoligotyping, *M. tuberculosis* complex strains were selected from an RIVM culture collection comprising about 8,000 clinical isolates collected between 1993 and 1998. All strains in this collection were typed by IS6110 fingerprinting, and those with less than five IS6110 copies had been subtyped by polymorphic GC-rich sequence

(PGRS) typing (39). A part of the strain collection (ca. 1,000 strains) has been spoligotyped as well. All *M. tuberculosis* complex strains, except the *M. bovis* BCG strains selected for hybridization with the novel spacers derived in this study differed in IS6110 restriction fragment length polymorphism (RFLP) and/or PGRS RFLP.

**Spoligotyping.** Culture and DNA isolation from mycobacteria was done as described previously (39). Spoligotyping using the previously published spacer oligonucleotides was done as described earlier (18). The incubation conditions of spoligotyping using the novel spacer oligonucleotides were slightly modified: hybridization was done for 45 min at 45°C, followed by two posthybridization washes for 10 min at 50°C and an incubation with streptavidin-peroxidase for 30 min at 42°C. All other conditions were as described by Kamerbeek et al. (18). Spoligotype designations were assigned to any unique hybridization pattern with the 43 different spacers used in traditional spoligotyping (18).

**DNA sequence analysis.** The published sequences of the DR region of *M. tuberculosis* strains 8 to 14, *M. bovis* strain 22, and the partial sequence of strain 20, the BCG Pasteur strain P3, were used in this study (see Table 1). The DR region sequences of all other strains listed in Table 1 were determined in this study. Sequencing was done by using the fluorescence-labeled dideoxy nucleotide technology using PCR products obtained by amplifying 0.1- to 3.0-kb fragments of the DR region, purified with Qiaquick purification kits (Qiagen, Hilden, Germany). For this purpose primer oligonucleotides were used based on sequences of DNA flanking the DR region in strain H37Rv (5), known spacer sequences, and sequences of the IS6110 element. DNA sequencing was done using an Applied Biosystems Instruments automatic sequencers (Models 373 and 377; Perkin-Elmer, Applied Biosystems Division). The sequences obtained were assembled, edited, and analyzed with the DNASTar package (DNASTar, Inc., Madison, Wis.). The sequences derived in this study were deposited at GenBank; for the accession numbers, see Table 1. Spacers were numbered according to their location in the genome (Fig. 1). Because this numbering differs from the previously assigned numbers in the study of Kamerbeek et al. (18), the old and the new spacer designations are given in Table 2.

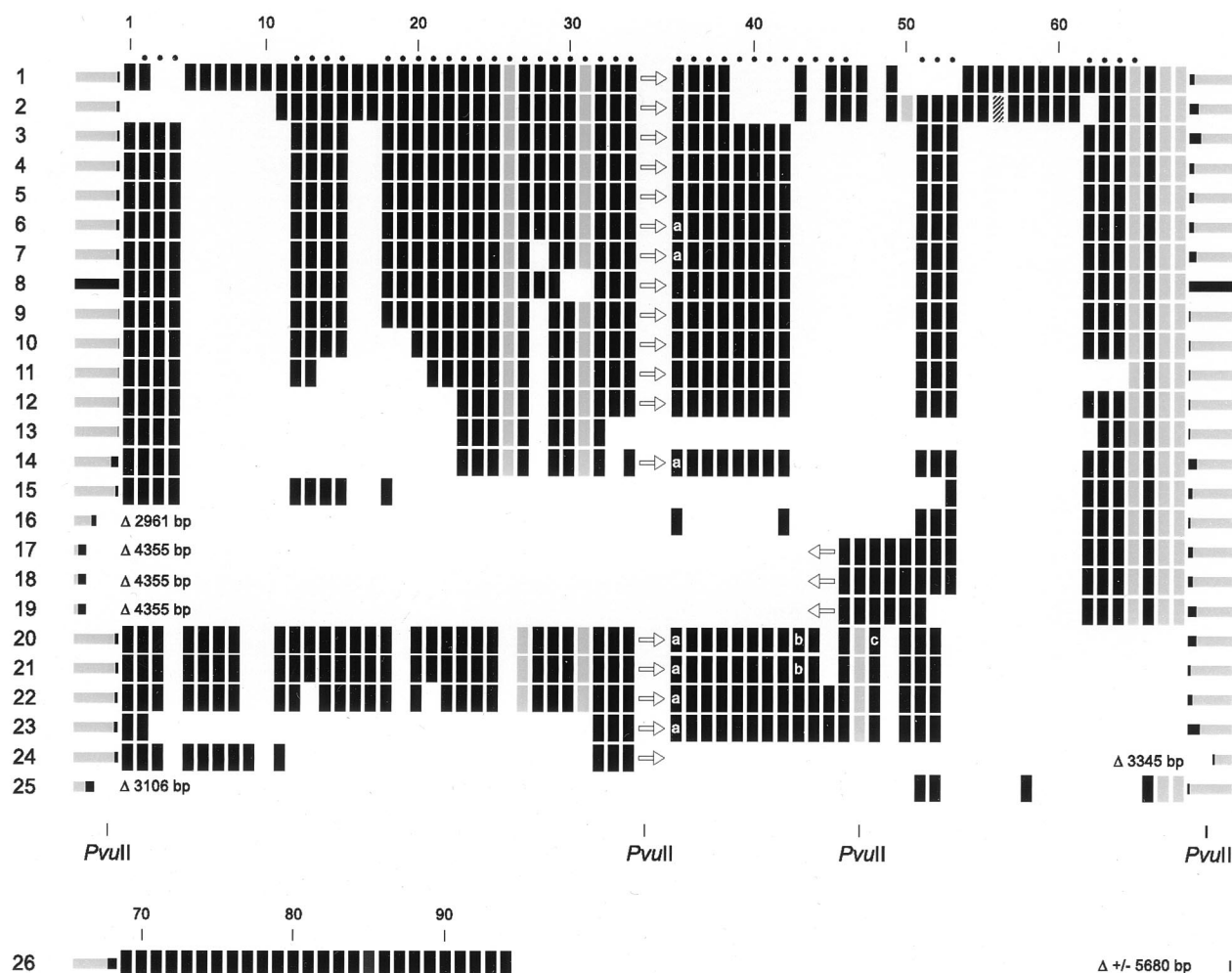


FIG. 1. Genetic organization of the DR locus in 26 *M. tuberculosis* complex strains as follows: 1 to 19, *M. tuberculosis*; 20 to 23, *M. bovis*; 24, intermediate phenotype between *M. tuberculosis* and *M. bovis*; 25, *M. microti*; and 26, *M. canettii*. The rectangles depict individual DVRs, which are composed of a DR and the adjacent spacer. Except for the *M. canettii* strain, vertically aligned rectangles represent DVRs with identical spacers. The sequence of the DR part of the rectangles is identical, except for those in gray. These differ in one or a few nucleotides from the consensus sequence. The hatched spacer in strain 2 differs in a single nucleotide from that in strain 1. The numbers at the top correspond to the spacer numbers listed in Table 2. The presence and the orientation of IS6110 is depicted by an arrow. DNA flanking the DR region is depicted by the bars at the left and at the right. The black parts of these bars depict the stretches that have been sequenced. The size of the DNA sequence missing compared to strain 8 (*M. tuberculosis* H37Rv) is given in base pairs, after the triangle. DVRs occurring twice in the DR region are depicted as rectangles marked with a letter: a, a duplication of DVR 35 is present 3' from spacer 45; b and c, duplicated DVRs 43 and 48, respectively, are present as tandem duplications.

## RESULTS

**DNA sequence analysis of the DR locus in *M. tuberculosis*, *M. bovis*, *M. microti*, and *M. canettii*.** The DNA sequence of the complete DR region was determined for 12 *M. tuberculosis* strains, 2 *M. bovis* strains, 2 *M. bovis* BCG strains, 1 *M. microti*, and 1 *M. canettii* strain. Furthermore, we sequenced a few hundred base pairs of the DNA flanking the DR region on either side. The results are summarized in Fig. 1, together with the previously published sequences of seven *M. tuberculosis* strains and one *M. bovis* strain. The size of the DR region varied from six DVRs for *M. microti* 19 to 56 DVRs for *M. tuberculosis* 1. Except for three spacers, all spacers were found only once in the DR locus. Spacer 35 was duplicated as a part of a DVR in all four *M. bovis* strains (strains 20 to 23) and in three *M. tuberculosis* strains (strains 6, 7, and 14). This duplicated spacer was separated from spacer 35 by six DVRs and was invariably located adjacent to spacer 41 (see Fig. 1). The two other duplicated spacers (spacers 43 and 49) were present as tandemly duplicated sequences. Spacer 43 was duplicated in

both *M. bovis* BCG strains (strains 20 and 21), and spacer 49 was duplicated in one BCG strain (strain 20). Virtually no interstrain variation in the sequences of spacers was observed. Only one spacer, spacer 56, present in *M. tuberculosis* strain 2 differed by a single nucleotide from the corresponding spacer in strain 1 (depicted by a hatched rectangle in Fig. 1; see also Table 2). A small degree of intrastrain sequence variation was found among the DRs. In six DVRs we encountered DRs with single-base-pair mismatches from the consensus sequence. Furthermore, the DR in DVR26 lacked four residues (Table 3). Probably DVR26 has been deleted because the adjacent spacer (spacer 25) is only 25 bp long, which is shorter than any other DVR (see Table 2). Interestingly, in all strains carrying mutant DRs, these mutant DRs invariably were present at the same location within the DR region.

For simplicity we numbered the DVRs and spacers according to their position in the DR locus, starting with number one at the left. Ninety-four different spacers were found among the 26 strains. Thirty-seven of these were not reported previously.

TABLE 2. List of spacer sequences in *M. tuberculosis* complex

Spacer no.	Previous designation (ref. 18)	Spacer sequence <sup>a)</sup>	Spacer no.	Previous designation (ref. 18)	Spacer sequence <sup>a)</sup>
1		TTAAACCGTGTGCACT <b>CAACCGGAATCTTGCA</b> C	48		GCGAGGA <b>ACCCTCCCACTGG</b> CGCTGCCCGACGCG
2 <sup>b)</sup>	1	CATAGAGGGT <b>CGCCGGCTCTGGATC</b> CGCTCCCTAGTCGT	49		TCAAT <b>TAACACTTTTTT</b> TTGAGCGTGGCGGTTGAGAGT
3 <sup>c)</sup>	2	TTTTTGCTCATGCTTGGGCGACAGCTTTTGCCAA	50		ACGGAA <b>ACGACAGCA</b> CGCTGACAATCTTATCTCGC
4	3	TCGCAAGCGCGGTCTTCCAGTGATCGCCTTCTA	51	37	ATTTT <b>GAGCGCAACTCGTCCACAGTCCCGCTT</b> TCAG
5		TCGCGCGCGGATGGCAGCG <b>AGCGGTGGCTAGGG</b>	52	38	GCCCC <b>GTGGATGGCGGATGCGTGTGCGG</b> CAAGT
6		ATGTGCGCGT <b>CGCCGTAACTGCC</b> CACGGCCCGT	53	39	CCGACGATGGCGAGTAAATCGCGGTGGTAACCGATCCGG
7		ATTTGACGACAAATTG <b>TTGACACGAATTTTCA</b> G	54		CCTCGCAGAAAAGGCGATCGATCATGAGAGTTGCGTTGAT
8		ACATCCACGCGTTACCGTGGCGCGCATATTCATCGA	55		AATGCTGGCGAGAT <b>TTTCGCTGTGTGGTTC</b> ATT
9		CCATATCGGGGACCGCGACGCTGCGAGAGGACACGCCA	56 <sup>d)</sup>		GCACAC <b>AGCACTCCCTTGACAT</b> CCGCGAGATCAGAC
10		TACACCACGCGT <b>CGTCCATCAGT</b> CAGCGCTCCTCCTC	57		TCGCGGGCTCGGCTAAGG <b>GTGCTGACTTCGCCTGTA</b>
11		TTGAACACGGAGCGCTGCA <b>ATGCCGTG</b> CTCAGGGGT	58		CCGACGAGAGCGAGCGCATACAGGCCAGGATACGCCAG
12	4	AACACCTCAGTAGAC <b>CTCATACGCCGACCAATCATAG</b>	59		TTG <b>CATCCACTGTGCG</b> CGACACGGCGACTTCCGCGA
13	5	TTTTCTGACCACTTGTGCGGGATTAGCGGGCTTAG	60		TGGTAAT <b>TGCTACGG</b> CGCGCTTGGCGGGCGATT
14	6	ACCAATCGCTCAT <b>TTCCGGCTTCAAT</b> TCAGCCT	61		ACCATCCGACGAGCGACCGAAGTCGATGACAAGCC
15	7	CTGAGGAGAGCGAGTACTCGGGCTCGCTGCTCGCGT	62	40	TAGTACGCCAT <b>CTGTGCTCTATACAGGT</b> CACTGCGCT
16		ACGACGT <b>TAGGCAATGCA</b> CTGCGCTCCCGCTTTTGA	63	41	CTGACGGCAGGAGCT <b>TTCCGGCTTCTATCAGGT</b> A
17		TGCTCTGAGCAACCGCATCATCCGGCGCGCGAGCTCCGC	64	42	CCTCATGTTGGGAGCATGAGCGAGCGCATATCGGG
18	8	CGGTGA <b>AAACCGCCCGCGCA</b> CTCGCGCGCGCTAG	65	43	TGGACCGAGAA <b>TGCACTCGCGGTGCGGAGT</b> CCAGCA
19	9	ACTCGGAATCCATGCTG <b>AGCAGCG</b> ATTTCGCAT	66		GCATATCGCCCGCACACACAG <b>CCACGCTACTGCTTCAT</b>
20	10	CGGGCAGCGTTCGACACCGCTCTA <b>GTGACTTCCG</b>	67		ACACCGCGGATGAGCA <b>TATGTCGAGT</b> GACATCCTCCCA
21	11	AGGTGAGCAACCGCGCGCGCACTCGCGCGCGGGTCTG	68		TTGAACCGCCCTTCGCGCT <b>GTGTTTCCGCGTGC</b> CGA
22	12	ATGGGATATCTGCTGCGCGCGCGGGAGATGCTGCTCGAG	69		TTGGAGGTGTG <b>GTGCAATCGCTGCACTG</b> CTGTTATT
23	13	TTGCTGACCACTATTCGCAT <b>TCCTCTCC</b> ACGT	70		ATGCCCGCTGAGTGGCGCCCGCGCGCGCGAGAA
24	14	TTGCGCAACCTTT <b>CGGTGTGATCGGATG</b> GTGCGCTCGG	71		TTGGTGATCT <b>CTCCATGACTTGACGCG</b> TTGACCGCAT
25	15	CTTGAATAACCGCGAGTGA <b>ATTTCCGGA</b>	72		GCGGTGCTGATCGCGCGCACTAGCGCGCGCAACTCGG
26	16	ATTTCGACGAGT <b>TTCCGTCAGCTC</b> GTAAATGCGCA	73		TCGGT <b>GCTGACCCATGGATG</b> GAGCGTAGACCTCTGT
27	17	CCGGCAACAATCGCGCGCGCGCGCGGATGACTCCG	74		CAACAAGGT <b>CTACGCGT</b> GAGGTCCACGGCTCAGAA
28	18	CGCATGGACCGCGCGCGAGCTCGAGATGCTCCGGGAG	75		ATGTGGCGATTACGCT <b>GTATCAGCGAAGGCGA</b>
29	19	TGGATTGCGCT <b>TAATGCTTGGCGT</b> GATCTCGGT	76		TTGATTAAT <b>TGACGCGACGCGCG</b> GAATCCAGCAACC
30	20	TCCACATCGATTTC <b>TTGACCTCGCCAGGAGAGATC</b> AC	77		CTTCAACGACCGCT <b>GTATTGGCCAT</b> GTGGGCTGGTCTTTCA
31	21	TCGTGACGATCGCGT <b>CGATGTGATGTC</b> CAATCGT <b>TCGA</b>	78		AGCAGCATGGACG <b>GTTCGCCTGTATCCGGGT</b> AAAAAA
32	22	TTGGAGCGTGT <b>CAACCGCAGCGCGCAGAT</b> TGAGACAA	79		TCGATGGCGCGCGGT <b>TGCGAATGTGGTGTG</b> CGGTAGC
33	23	CCTCAGCT <b>CAGCATCGCTGATCGGTC</b> AGCTCTCCCT	80		TTGGCGT <b>ACATAGCGAGCTG</b> TGGCGCAGTAGAGTGGCG
34	24	CCAACCTCACCGCTGCTGGGTGAGACGTGCTC <b>CGCGCA</b>	81		CTTCGATTGTGCGCGCGCGGT <b>TTCTGTTAC</b>
35	25	TCGGGGAGCGGAT <b>CAGCGACACCGCA</b> CCCTGTCA	82		TCGCCGGCT <b>TGGCGCTGTGTTCTGAT</b> CTCGCTAAT
36	26	CTT <b>CAGCACCACCATCAT</b> CCCGCGCTCAGCTCAGCAT	83		CCTTGCCAAAGGAGGT <b>CTGCTGTGCTTATGCT</b> ACAG
37	27	CCTTCGACGCGGAT <b>TCGTGATCTCTTCCCGCGGATG</b>	84		TAGTTGACGCGAAAT <b>GTGTTGAGCTGTGATCA</b> ATTCAA
38	28	TGCGCGCGGT <b>TTAGCGATCAACA</b> CCCACTAATG	85		GCTTGACAATGCGGT <b>TTTCGCGCGCTTTTCCAGCCGAG</b>
39	29	CAGCGAAATACAGGCT <b>CCACGACACGAC</b> CAACGC	86		GTGTGTTTCAACG <b>GTTCAGTTTCTTGTCCCA</b> GTG
40	30	TCCTTGACGATGCGGT <b>TGCGCCCGCGCTTTTCCAGCC</b>	87		ACTGGT <b>GTGTTGCCCGG</b> GACGCGCGGTGGAGTGTG
41	31	AGGTTGCGCT <b>TCAGACAGGTTCGCGT</b> CGATCAAGT <b>TCG</b>	88		CATCCAGAGGT <b>CGAAGTGGTGTGCGTGTTCCT</b> GTIAC
42	32	TTTATCACTCCG <b>ACCAATAGGTATCGCGCTGTTC</b> AA	89		TTCTGCGCGCGATCCGGT <b>CTATGACGAGCCGCGAG</b>
43	33	TCGACACCGAT <b>ATGACGCGGTGCGCGACTT</b> GCAGCA	90		ATCAG <b>CACAGCGCTT</b> AGCGTTCGGTGGCGGCAAGTCTCAGA
44	34	CTTTGCGAAGT <b>CACCTCGCCACACCGT</b> CGAAGCGCCT	91		CAGAAGGTAGCGTCCGCGGAT <b>TGTCTGCGCCAC</b>
45		GCGGATGTTGGGCA <b>AGTTGGCGCTGGGGT</b> CTGAGTCAA	92		CATCGAACAAGCGCGCTCGG <b>TAAACAGCGCTGTCTCA</b>
46	35	TCGCATCCGGAAAT <b>CTCGTACGCTCGA</b> AACTTCAAGCT	93		TCTCCGTGATAGGTGAG <b>ACACCGCAATCACCATCA</b>
47	36	TCGAAATCCAGCACCAT <b>CCGACG</b> TCGCGCATGCTCCCGAA	94		TCTGGT <b>AGTGGCGTCT</b> CGCGGTGCGCGGTGTGTCT

<sup>a</sup> Bold sequences used for oligonucleotide probes in spoligotyping.<sup>b</sup> The previously published sequence (18) contained a C at residue 17 erroneously.<sup>c</sup> The previously published sequence (18) contained two errors at residues 13 and 14 (both were A).<sup>d</sup> This spacer is polymorphic at position 21. Strain 1 is as shown, and strain 2 has a C at this position.

Twenty-six of the novel spacers (spacers 69 to 94) were present only in *M. canettii*. The remaining 11 novel spacers were all found in *M. tuberculosis* 1, and nine of these spacers were present in only one other strain, *M. tuberculosis* 2. *M. tuberculosis* strains 1 and 2 are exceptional because of the presence of only a single copy of IS6110. The four *M. bovis* strains investigated differed from all *M. tuberculosis* strains in the presence of spacer 44 and in the absence of 16 spacers (spacers 53 and 54 to 68). These data confirm the previously observed *M. bovis*-characteristic signature in the DR region (18). The most remarkable finding is the strong conservation of the order of the DVRs in the various isolates. As depicted in Fig. 1, the polymorphism in the DR region appears to comprise mainly the presence or absence of single, discrete DVRs or stretches of contiguous DVRs. Except for the duplication of DVR35, no scrambling of DVRs appears to have taken place during the evolution of the DR region.

The DNA flanking the DR locus in most strains was identical to that in *M. tuberculosis* H37Rv (strain 8 in Fig. 1). Seven

strains lacked 3- to 6-kb stretches 5' or 3' of the DR region. In all seven strains lacking DR-flanking DNA, IS6110 was either directly adjacent to the flanking DNA or IS6110 was absent from the DR region, suggesting the involvement of this insertion element in the deletion of chromosomal DNA. The DNA

TABLE 3. Variation in DR sequences

DVR type	Sequence <sup>a</sup>
Consensus sequence	.....GTCGTCAGACCCAAACCCGAGAGGGGACGAAAC
DVR26	-----TCAGACCCAAACCCGAGAGGGGACGAAAC
DVR31	.....GTCGTCAGACCCAAACCCGAGAGGGGACGAAAC
DVR47	.....GTCGTCAGACCCAAACCCGAGAGGGGACGTAAC
DVR50	.....GTCGTCAGACCTAAACCCGAGAGGGGACGAAAC
DVR65	.....GTCGTCGAGACCCAAACCCGAGAGGGGACGAAAC
DVR67	.....GTCCTCAGACCCAAACCCGAGAGGGGACGAAAC
DVR68	.....GTCGTCAGACCCAAACCCGAGAGGGGACGAAAC

<sup>a</sup> Polymorphisms are indicated in boldface.

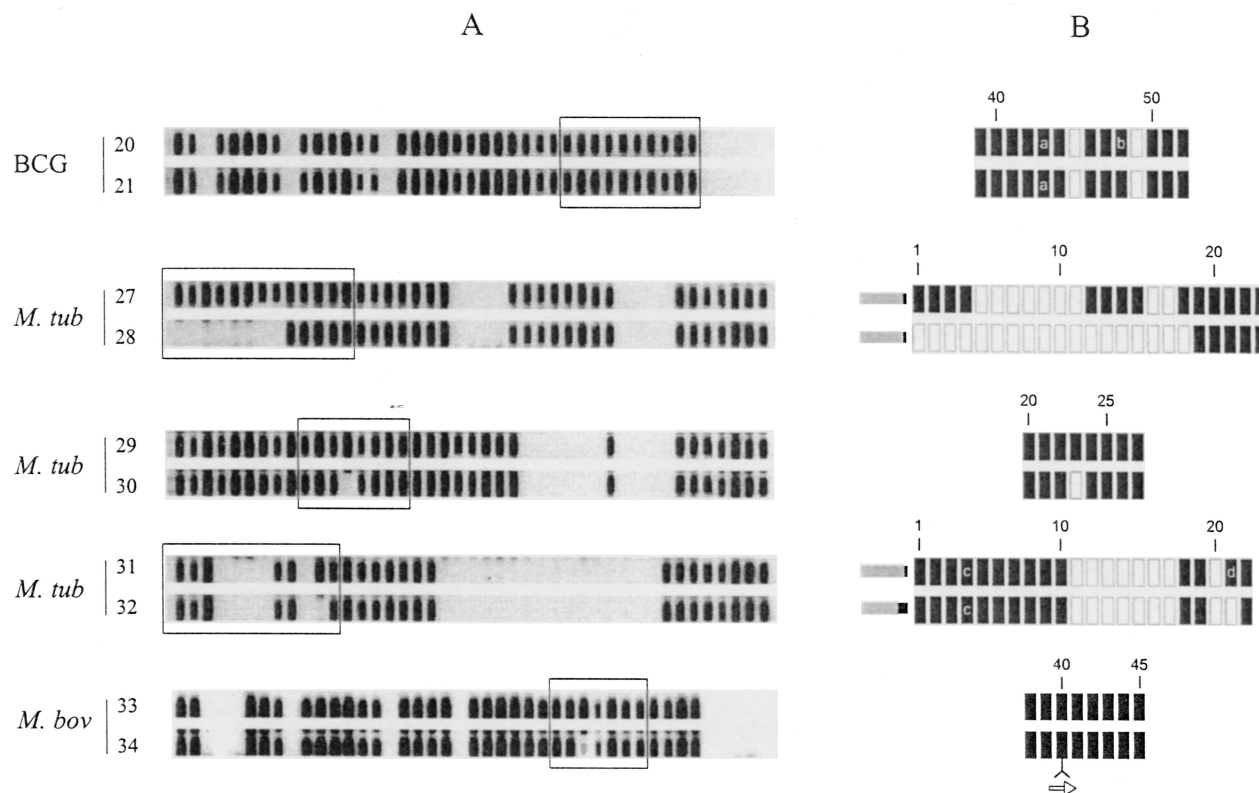


FIG. 2. Rearrangements in the DR locus of presumed isogenic variants. (A) Spoligotype of each of the five isogenic pairs determined by using the standard 43 spacers as a probe; the boxed areas have been sequenced and are depicted in panel B. (B) Arrangement of DVRs as determined by DNA sequencing; the numbers correspond to the spacer numbers as given in Fig. 1. Black rectangles depict DVRs that are present; gray rectangles represent DVRs that are absent. Numbers correspond to DVR numbering as shown in Fig. 1. The arrow represents the insertion element IS6110. Strain numbers are depicted on the left. a, b, and d, tandem duplications of DVR43; DVR48, and DVR21, respectively; c, a duplicated copy of DVR4 is located directly to the right of DVR19.

flanking the 3' side of the DR region in the *M. canettii* strain shared no similarity with the sequence of H37Rv. Interestingly, this DNA was found to share 80% identity to the insertion element IS1096 from *M. smegmatis* (25). Strain 24 was the only other strain lacking sequences in the 3'-flanking region. This strain could not be classified as any of the known *M. tuberculosis* complex species because it shared biochemical characteristics of both *M. tuberculosis* and *M. bovis*. This may suggest that this strain belongs to a distinct, as-yet-unrecognized taxon within the *M. tuberculosis* complex group of bacteria. Because the origin of DRs and spacers in the DR region is unknown, we checked the GenBank database for sequences homologous to the DR and all of the spacers. No significant sequence similarity was found.

**Rearrangements in the DR locus of presumed isogenic variants.** To investigate the nature of the rearrangements in the DR region, we attempted to collect isogenic strain pairs that differ in the DR locus. Only a single set of true isogenic strains was available. This set comprised the two *M. bovis* BCG strains investigated in this study. The Russian BCG strain, strain 21, differed from the Pasteur vaccine strain, strain 20, in the absence of a duplication of DVR49 as described above (Fig. 2).

Unfortunately, we have been unable to derive in the laboratory subcultures which differ from parental strains in the DR region. Therefore, we have analyzed pairs of strains which are very likely to be isogenic, either because they were isolated during a single outbreak or because they originated from the same patient. The results of partial sequencing of the DR region of these strains are depicted in Fig. 2. *M. tuberculosis*

strains 27 and 28 were isolated from two different body sites of a 76-year-old tuberculosis patient, who probably reactivated from an infection acquired early in life. The IS6110 RFLP patterns of these isolates are related because 10 *Pvu*II IS6110-containing restriction fragments are shared among the 12 IS6110 fragments in strain 27 and the 14 fragments in strain 28 (41). We assume that early in life the patient has been infected with a single strain and that the long period of dormancy allowed DNA rearrangements to take place. Strain 27 differed from strain 28 by the absence of 8 spacers in spoligotyping (Fig. 2A). Sequencing showed these two that strains differ in the absence of exactly nine discrete DVRs, of which eight are used in traditional spoligotyping using the set of 43 spacer probes (Fig. 2B). The third pair of strains comprised *M. tuberculosis* strains 29 and 30, isolated from a 73-year-old Dutch patient. These strains have identical IS6110 RFLP patterns (showing the presence of eight IS6110 copies) but differ in a single spacer in spoligotyping. Sequencing showed that strain 30 differed from strain 29 in the absence of a single DVR, DVR23. The fourth presumed isogenic pair comprised *M. tuberculosis* strains 31 and 32, isolated from two Somali immigrants, who probably were epidemiologically linked, although this link could not be confirmed by traditional contact tracing. Both strains have identical IS6110 RFLP patterns (showing 13 IS6110 copies) but differed in a single spacer reaction by spoligotyping. Sequencing showed that strain 31 differed from strain 32 in the presence of two DVRs, which are tandem duplications of DVR21. Finally, we investigated *M. bovis* strains 33 and 34, isolated during an epidemic of multidrug-resistant tuberculosis

in Spain (3). These two strains differ in a single IS6110-containing restriction fragment and in a single spacer reaction by spoligotyping (Fig. 2A). Sequencing showed that strain 34 differed from 33 in the presence of a copy of IS6110 in the DR of DVR40 (Fig. 2B). We conclude that this insertion prevented amplification of DVR40 during the PCR and thus resulted in the apparent absence in the spoligotyping pattern.

Although there is no direct evidence that four of the five strain pairs investigated are true isogenic pairs, the nature of the genetic variation found in the DR loci is consistent with this assumption. In four of the strain pairs the genetic rearrangements can be explained by a single genetic event comprising either the insertion or the deletion of a single, discrete DVR or a set of contiguous DVRs. In the remaining pair the variation is also explained by a single genetic event, namely, the insertion or deletion of IS6110 in the DR region.

**Strain differentiation of *M. tuberculosis* complex strains by using novel spacers.** We investigated whether the degree of strain differentiation would be improved by typing using more spacers than the 43 spacers used in standard spoligotyping as described by Kamerbeek et al. (18). For this purpose we spoligotyped *M. tuberculosis* complex strains using the standard set of 43 spacers plus the 51 novel spacers. We analyzed 170 clinical isolates of *M. tuberculosis* complex. All of these isolates had previously been fingerprinted by IS6110 and, when fewer than five IS6110 copies were present, the PGRS subtypes were also known. Group 1 comprised 65 *M. tuberculosis* strains harboring a single IS6110 copy or only two IS6110 copies. Such strains are known to be hard to differentiate by IS6110 fingerprinting (32). By traditional spoligotyping these strains were differentiated into nine spoligotypes. By using the novel spacers, 23 strains were further subtyped, and the number of different hybridization patterns increased to 20 (Fig. 3). Interestingly, the group 1 strains displayed a characteristic spacer signature in the region of spacer 37 to 48 (Fig. 3).

Group 2 strains comprised 26 *M. tuberculosis* strains that were selected because their spoligotypes corresponded to the most frequently encountered four spoligotypes among patients in The Netherlands. None of these strains were subdivided with the novel spacer set (Fig. 3).

Group 3 strains comprised 35 *M. tuberculosis* strains, which hybridized with relatively few spacers from the old set of 43 in standard spoligotyping. Fifteen of these strains were of the spoligotype, which is characteristic for strains of the Beijing genotype (38). By using the novel spacers only a single strain among the 35 strains tested was further differentiated (Fig. 3). Therefore, the use of additional spacers in spoligotyping led to a marginal increase in the level of strain differentiation. Compared to spoligotyping, IS6110 fingerprinting differentiated better for the strains harboring multiple IS6110 copies, and among the low-copy strains spoligotyping differentiated the strains slightly better.

A group of 38 *M. bovis* strains was investigated. *M. bovis* usually harbors only one or a few copies of IS6110, and therefore such strains are notoriously difficult to differentiate (6). The most common spoligotypes among *M. bovis* isolated in The Netherlands are the types 120 and 121 (unpublished observations). Therefore, we included 34 strains with these two spoligotypes for testing with the additional 51 novel spacer probes (Fig. 3). Seven *M. bovis* strains were subdivided, and the number of types increased from four to eight. Thus, similar to the low-IS6110-copy *M. tuberculosis* strains, the novel spacers contributed significantly to improving the strain differentiation of *M. bovis*.

The two *M. microti* strains shared the six spacers characterized by sequencing one strain. The other one contained, in ad-

dition, 15 other spacers (Fig. 3). Finally, we investigated three *M. canettii* strains, isolated in France, The Netherlands, and Switzerland (28, 37). All of these strains displayed a hybridization pattern identical to that of the 94 spacer probes. These strains did not share any spacers with those found in the other *M. tuberculosis* complex strains (Fig. 3).

## DISCUSSION

In this study we compared the sequence of the complete DR locus and the bordering DNA in 26 *M. tuberculosis* complex strains in an attempt to better understand the mechanisms underlying the genesis and evolution of these peculiar genetic elements with unknown function. The size of the DR locus varied from 6 DVRs (0.6 kb) to 56 DVRs (6 kb), and both spacers and DRs showed little sequence interstrain variation. The most remarkable finding is the strong conservation of the order of the various DVRs among the strains. We did not find a single strain in which the order of the DVRs differed. This indicates that during evolution individual DVRs did not move within the DR region. In addition, the individual DRs also do not seem to move because the six single DRs, which varied slightly from the consensus sequence, were all found at the same position within the DR region. Except for occasional insertion-element-driven polymorphisms, we found that the major type of polymorphism comprised the presence or absence of single, discrete DVRs or stretches of contiguous DVRs. The data suggest that the DR loci in the clinical isolates we see today are the remnant of a primordial DR locus, which was composed of a large number, perhaps hundreds, of different DVRs. The most likely mechanism underlying the strain-to-strain variation is the successive deletion of single or multiple discrete DVRs from this archetypal DR region. We also observed a few rare duplications of discrete DVRs, mostly as tandem duplications. These deletions and duplications probably have been mediated by homologous recombination between neighboring or distant DRs and/or by slippage during DNA replication. The rearrangements observed in presumed isogenic strains are consistent with this view. These strains differed in the presence or absence of a single discrete DVR or a stretch of contiguous DVRs and the presence or absence of IS6110 at an uncommon site in the DR region. These observations are consistent with the evolutionary scenario recently proposed by Fang et al. (10) for a closely related group of *M. tuberculosis* strains isolated from different geographic areas. In this scenario strain variation was thought to be due to the deletion of discrete DVRs in the DR region and due to the transposition of IS6110.

In some of the strains we observed deletions in the DNA flanking the DR region. In these strains DVRs were missing DVRs, which in other strains were present at the left or the right border of the DR locus (see Fig. 1). This indicates that the deletions in DR flanks took place concurrently with deletion of the DVRs, which normally delimit the DR region. The lack of the IS6110 element in these strains strongly suggests that these deletions were IS6110 mediated. These observations are consistent with the idea of a primordial DR locus in which successive deletions led to the presently observed genetic variation.

Although variation by deletion from a primordial DR locus seems the easiest explanation for the genetic variation in the DR locus of present-day strains, the genesis of the hypothetical primordial DR locus remains enigmatic. Presently, the complete genome sequence of two *M. tuberculosis* strains is known. We have searched these genomes for sequence similarity with any of the 94 different spacer sequences, but except for se-

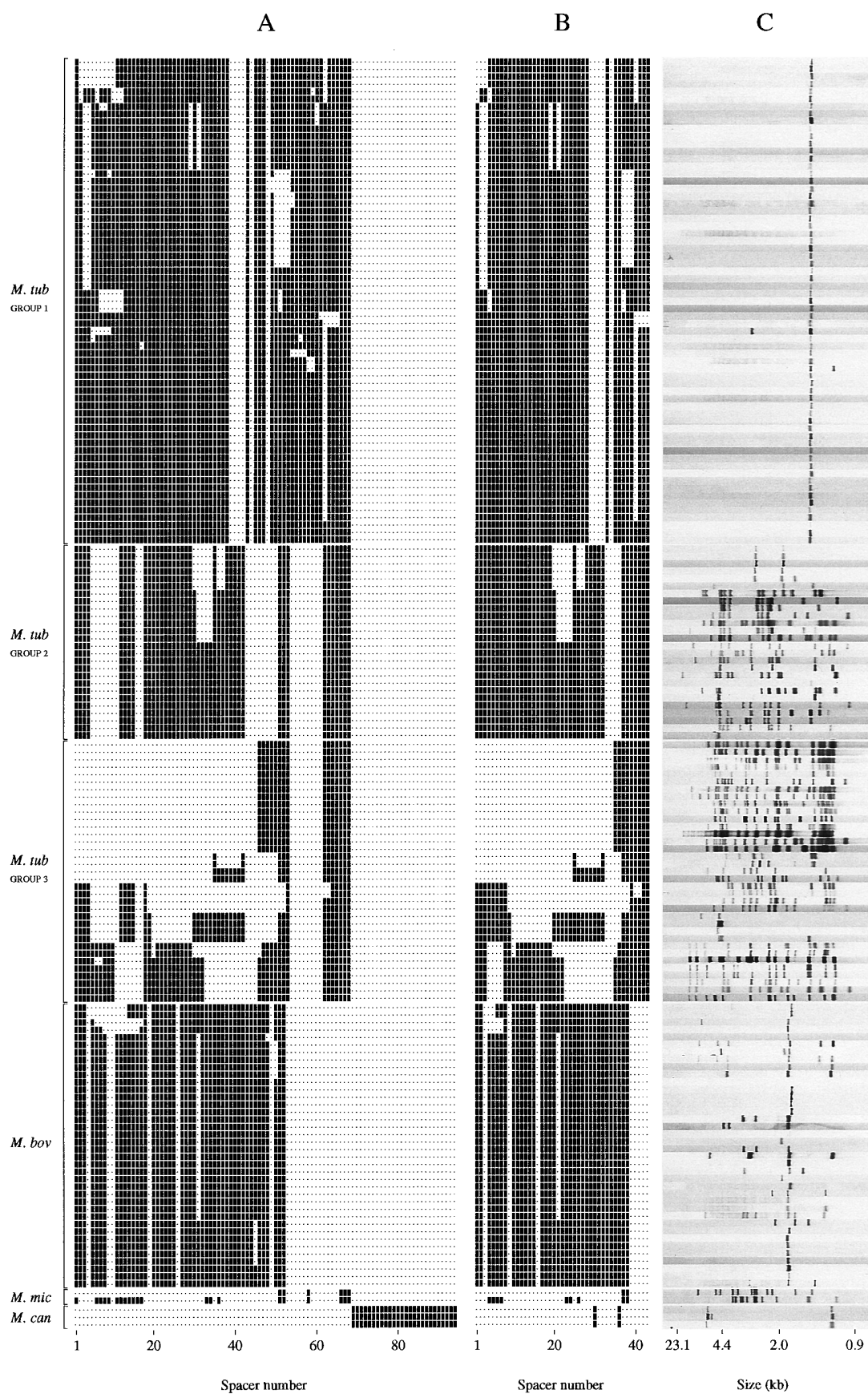


FIG. 3. Presence of spacers in and traditional spoligotypes and IS6110 RFLP patterns of 170 *M. tuberculosis* complex strains. (A) Presence or absence of any of the 94 spacers; the spacers are ordered as in Fig. 1, and this order corresponds to the presumed order in the genome. (B) Hybridization signals of the 43 spacers used in traditional spoligotyping. (C) IS6110 RFLP patterns. Strains of group 1 belong to the commonly found spoligotypes ST5, ST14, ST22, and ST38.

quences within the DR locus no significant sequence similarity was found. Therefore, it seems unlikely that in the present-day strains novel spacers in the DR locus are generated from a template of existing sequences elsewhere in the mycobacterial genome. At one time the DR perhaps had the capacity to multiply by replicative transposition or retroposition within a nonessential region of the genome. However, at present no examples in nature are known by which short pieces of DNA are duplicated in such a way that the repeats become separated by similarly sized nonrepetitive intervening sequences. Perhaps the DVRs evolved from directly repeated DNA without intervening spacer DNA. These repeats could have acquired a biological function, such as replicon partitioning, as has been found in *Haloferax* spp. (26). When the selective force was imposed on repeat length and part of the specific repeat sequence, such repeats could have diverged to the present-day DVR elements with a constant part and a variable part. It should be noted that in bacterial plasmids repeats have been identified which are involved in the regulation of replication and plasmid compatibility. These repeats or "iterons" have sizes similar to those of the DVRs, and also some sequence variation is found within the iterons on a single plasmid (7). The number of iterons per cell determines the copy number and stability. Perhaps DVRs have evolved from repeats with a similar function as iterons in plasmids.

Previous studies have revealed the existence of 57 different spacers. Forty-three of these have been used in standard spoligotyping for strain differentiation of *M. tuberculosis* complex isolates on the basis of the strain-dependent presence or absence of these spacers (18). In this study we disclosed 37 novel spacer sequences. The majority of these, 26 sequences, were found in *M. canettii*, a recently described taxon within the *M. tuberculosis* complex group of bacteria (37). *M. canettii* shared not a single spacer with other *M. tuberculosis* complex strains, and none of the *M. canettii* spacers were found in other *M. tuberculosis* complex strains. Thus, the DR locus in *M. canettii* differs greatly from the DR loci in other members of the *M. tuberculosis* complex. This difference confirms previous observations showing that *M. canettii* differs in many respects from the other species in this group of mycobacteria, such as the presence of multiple mutations in certain housekeeping genes, multiple chromosomal deletions, and differences in the cell wall composition (37). Other than the large number of *M. canettii*-specific spacers, we disclosed 11 novel spacer sequences in *M. tuberculosis* and *M. bovis*. The use of these novel spacers for strain typing improved the degree strain differentiation, in particular of strains harboring few IS6110 copies. From the very first publications on spoligotyping it has appeared that strains with certain spoligotypes are polymorphic when analyzed by other genetic markers such as IS6110 (13, 18, 22). In this study we have confirmed that strains with certain common spoligotypes encountered among clinical isolates of *M. tuberculosis* have an identical or almost identical DR region sequence in spite of their unrelated IS6110 and PGRS RFLP patterns. This indicates that the DR region remained unchanged during a long period of time during which genetic rearrangements took place at other chromosomal loci. Three mutually nonexclusive explanations are possible: (i) the DNA arrangement of the DR region in these strains is frozen because of an unknown structural property of the specific DR region sequence or because of a poor ability in homologous recombination or slipped strand mispairing during replication; (ii) the specific sequence of the DR region in these strains provides them with a selective advantage and therefore variants with DVR rearrangements do not persist in the population, or (iii) these strains acquired the DR region from other

strains by horizontal DNA transfer. The latter possibility seems unlikely because other studies suggest that the population structure of *M. tuberculosis* is clonal rather than panmictic (11, 22). Furthermore, certain spacers in *M. bovis* and *M. canettii* were not found in *M. tuberculosis*, suggesting the absence of lateral transfer of DVR sequences. Presently, it is impossible to distinguish between the first two possibilities because the function of the DR region in *M. tuberculosis* is unknown and we have not been able to derive *M. tuberculosis* mutants with a rearranged DR region in the laboratory. The only bacterial species in which a function of a DR-like region has been proposed is *Haloferax* spp., in which the number of repeats seems to be involved in replicon partitioning (26).

In this study we confirmed the existence *M. tuberculosis* strains which are genetically divergent as measured with markers such as IS6110 and PGRS and which even so have an identical DR region. This observation complicates the use of spoligotyping for epidemiological analysis. Ideally, polymorphic genetic markers should have molecular clocks with equal paces for different strains. Our study strongly suggests that large differences among *M. tuberculosis* strains exist in the pace of the molecular clock of the DR region. Strains with spoligotypes such as type 38 often exhibit very different IS6110 RFLP patterns, although their spoligotypes are identical, indicating that molecular changes due to the mobility of IS6110 is quicker than that of rearrangements in the DR locus. We found the reverse situation in strains with a single or few IS6110 copies. In such strains the insertion element seems to be "frozen," whereas they do exhibit DR-associated DNA polymorphism. The apparent strain-dependent pace of different molecular clocks of the various genetic markers used in the epidemiology of tuberculosis underlines the point that great care should be taken in the interpretation of strain typing in particular when only a single genetic marker is used.

#### ACKNOWLEDGMENTS

This work was financially supported by the Dutch Foundation for Technical Sciences and the European Union project on the development of novel standardized methodology and nomenclature for the identification of *M. bovis* strains.

We acknowledge Marjori Beggs, David Brittain, Solvig Roring, Robin Skuce, and Z. Fang for providing us with unpublished sequences of the DR region in *M. bovis* and *M. tuberculosis*.

#### REFERENCES

- Bachelier, S., E. Gilson, M. Hofnung, and C. W. Hill. 1996. Repeated sequences, p. 2012–2040. In F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, Jr., B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: American Society for cellular and molecular biology, 2nd ed. Microbiology, Washington, D.C.
- Beggs, M. L., M. D. Cave, C. Marlowe, L. Cloney, P. Duck, and K. D. Eisenach. 1996. Characterization of *Mycobacterium tuberculosis* complex direct repeat sequence for use in cycling probe reaction. *J. Clin. Microbiol.* **34**: 2985–2989.
- Blazquez, J., L. E. Espinoza de Los Monteros, S. Samper, C. Martin, A. Guerrero, J. Cobo, J. van Embden, F. Baquero, and E. Gomez-Mampaso. 1997. Genetic characterization of multidrug-resistant *Mycobacterium bovis* from a hospital outbreak involving human immunodeficiency virus-positive patients. *J. Clin. Microbiol.* **35**:1390–1393.
- Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. D. Scott, N. S. Geoghagen, J. F. Weidman, J. L. Fuhrmann, D. T. Nguyen, T. Utterback, J. M. Kelley, J. D. Peterson, P. W. Sadow, M. C. Hanna, M. D. Cotton, M. A. Hurst, K. M. Roberts, B. B. Kaine, M. Borodovsky, H. P. Klenk, C. M. Fraser, H. O. Smith, C. R. Woese, and J. C. Venter. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**:1058–1073.
- Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V.

- Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, S. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, S. Skelton, S. Squares, R. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.
6. Cousins, D., S. Williams, E. Liebana, A. Aranaz, A. Bunschoten, J. D. A. van Embden, and T. Ellis. 1998. Evaluation of four DNA typing techniques in epidemiological investigations of bovine tuberculosis. *J. Clin. Microbiol.* **36**: 168–178.
7. DelSolar, G., R. Giraldo, M. J. RuizEchevarria, M. Espinosa, and R. DiazOrejas. 1998. Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**:434–464.
8. Diaz, R., K. Kremer, P. de Haas, R. I. Gomez, A. Marrero, J. A. Valdivia, J. D. A. van Embden, and D. van Soolingen. 1998. Molecular epidemiology of tuberculosis in Cuba outside of Havana, July 1994–June 1995: utility of spoligotyping versus IS6110 restriction fragment length polymorphism. *Int. J. Tuberc. Lung. Dis.* **2**:743–750.
9. Driscoll, J. R., M. A. McGarry, and H. W. Taber. 1998. DNA typing of a nonviable culture of *Mycobacterium tuberculosis* in a homeless shelter outbreak. *J. Clin. Microbiol.* **37**:274–275.
10. Fang, Z., N. Morrison, B. Watt, C. Doig, and K. J. Forbes. 1998. IS6110 transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. *J. Bacteriol.* **180**:2102–2109.
11. Frothingham, R., and W. A. Meeker-O'Connell. 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* **144**:1189–1196.
12. Gordon, S. V., R. Brosch, A. Billault, T. Garnier, K. Eiglmeijer, and S. T. Cole. 1999. Identification of variable regions in the genome of tubercle bacilli using artificial chromosome arrays. *Mol. Microbiol.* **32**:643–655.
13. Goyal, M., N. A. Saunders, J. D. A. van Embden, D. B. Young, and R. J. Shaw. 1997. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS6110 restriction fragment length polymorphism. *J. Clin. Microbiol.* **35**:647–651.
14. Groenen, P. M. A., A. E. Bunschoten, D. van Soolingen, and J. D. A. van Embden. 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*: application for strain differentiation by a novel typing method. *Mol. Microbiol.* **10**:1057–1085.
15. Gutiérrez, M., S. Samper, M. S. Jiménez, J. D. A. Embden, J. F. García-Marín, and C. Martín. 1997. Identification by spoligotyping of a caprine genotype in *Mycobacterium bovis* strains causing human tuberculosis. *J. Clin. Microbiol.* **35**:3328–3330.
16. Hermans, P. W. M., D. van Soolingen, E. M. Bik, P. E. W. de Haas, J. W. Dale, and J. D. A. van Embden. 1991. The insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect. Immun.* **59**:2695–2705.
17. Hoe, N., K. Nakashima, D. Grigsby, X. Pan, S. J. Dou, S. Naidich, M. Garcia, E. Kahn, D. Bergmire-Sweat, and J. M. Musser. 1999. Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. *Emerg. Infect. Dis.* **5**:254–263.
18. Kamerbeek, J., L. M. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, J. E. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. D. A. van Embden. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**:907–914.
19. Kapur, V., T. S. Whittam, and J. M. Musser. 1994. Is *Mycobacterium tuberculosis* 15,000 years old? *J. Infect. Dis.* **170**:1348–1349.
20. Kwarabiyasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, Y. Nakamura, T. F. Robb, K. Horikoshi, Y. Masuchi, H. Shizuya, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**:55–76.
21. Klenk, H. P., R. A. Clayton, J. F. Tomb, O. White, K. E. Nelson, K. A. Ketchum, R. J. Dodson, M. Gwinn, E. K. Hickey, J. D. Peterson, D. L. Richardson, A. R. Kerlavage, D. E. Graham, B. A. Kyrpides, R. D. Fleischmann, J. Quackenbush, N. H. Lee, G. Sutton, S. Gill, E. F. Kirkness, B. A. Dougherty, K. McKenney, M. D. Adams, B. Loftus, S. Peterson, C. I. Reich, L. K. McNeil, J. H. Badger, A. Glodek, L. X. Zhou, R. Overbeek, J. D. Gocayne, J. F. Weidman, L. McDonald, T. Utterback, M. D. Cotton, T. Spriggs, P. Artiach, B. P. Kaine, S. M. Sykes, P. W. Sadow, K. P. Dandrea, C. Bowman, C. Fujii, S. A. Garland, T. M. Mason, G. J. Olsen, C. M. Fraser, H. O. Smith, C. R. Woese, and J. C. Venter. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**:364–370.
22. Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. M. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. Riley, M. A. Yakus, J. M. Musser, and J. D. A. van Embden. 1999. Comparison of methods for typing of *Mycobacterium tuberculosis* complex strains for use in the epidemiology: an interlaboratory study on strain discriminatory power and reproducibility. *J. Clin. Microbiol.* **37**:2607–2618.
23. Long, R., E. Nobert, S. Chomyc, J. D. A. van Embden, C. McNamee, R. R. Duran, J. Talbot, and A. Fanning. 1999. Transcontinental spread of multi-drug-resistant *Mycobacterium bovis*. *Am. J. Respir. Crit. Care Med.* **159**: 2014–2017.
24. Masepohl, B., K. Gortitz, and H. Bohme. 1996. Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Biochim. Biophys. Acta Gene Struct. Expr.* **1307**:26–30.
25. McAdam, R. A., T. R. Weisbrod, J. Martin, J. D. Scuderi, A. M. Brown, J. D. Cirillo, B. Bloom, and W. R. Jacobs, Jr. 1995. In vivo growth characteristics of leucine and methionine auxotrophic mutants of *Mycobacterium bovis* BCG generated by transposon mutagenesis. *Infect. Immun.* **63**:1004–1012.
26. Mojica, F. J. M., G. Ferrer, and F. Rodriguez-Valera. 1995. Long stretches of short tandem repeats are present in the largest replicons of the archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* **17**:85–93.
27. Nakata, A., M. Amemura, and K. Makino. 1989. Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J. Bacteriol.* **171**:3353–3356.
28. Pfyffer, G., E. R. Auckenthaler, J. D. A. van Embden, and D. van Soolingen. 1998. *Mycobacterium canettii*, the smooth variant of *M. tuberculosis*, isolated from a Swiss patient exposed in Africa. *Emerg. Infect. Dis.* **4**:631–634.
29. Poulet, S., and S. T. Cole. 1995. Repeated DNA sequences in mycobacteria. *Arch. Microbiol.* **163**:79–86.
30. Qian, L., J. T. Douglas, A. van der Zanden, H. Duanmu, and J. D. A. van Embden. 1999. Retrospective analysis of *M. tuberculosis* DNA types in preserved lung tissues. *J. Clin. Microbiol.* **37**:471–474.
31. Shreevatsan, S., P. Escalante, X. Pan, K. E. Stockbauer, N. D. Conell, B. N. Kreiwirth, T. S. Whittam, and J. M. Musser. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionary recent global dissemination. *Proc. Natl. Acad. Sci. USA* **94**:9869–9874.
32. Small, P. M., and J. D. A. van Embden. 1994. Molecular epidemiology of tuberculosis, p. 569–582. In B. Bloom (ed.), *Tuberculosis: pathogenesis, protection, and control*. ASM Press, Washington, D.C.
33. Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, and J. N. Reeve. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol.* **179**:7135–7155.
34. Sola, C., A. Devallois, L. Horgen, J. Maisetti, I. Filliol, E. Legrand, and N. Rastogi. 1999. Tuberculosis in the Caribbean: using spacer oligonucleotide typing to understand strain origin and transmission. *Emerg. Infect. Dis.* **5**: 404–414.
35. Taylor, G. M., M. Goyal, A. J. Legge, R. J. Shaw, and D. Young. 1999. Genotypic analysis of *Mycobacterium tuberculosis* from medieval human remains. *Microbiology* **145**:899–904.
36. van der Zanden, A. G. M., A. H. Hoentjen, F. G. C. Heilmann, E. F. Weltevreden, L. M. Schouls, and J. D. A. van Embden. 1998. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* complex in paraffin wax embedded tissues and in stained microscopic preparations. *Mol. Pathol.* **51**:209–214.
37. van Soolingen, D., T. Hoogenboezem, P. E. W. de Haas, P. W. M. Hermans, M. A. Koedam, K. S. Teppema, P. J. Brennan, G. S. Besra, F. Portaels, J. Top, L. M. Schouls, and J. D. A. van Embden. 1997. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *Int. J. Syst. Bacteriol.* **47**:1236–1245.
38. van Soolingen, D., L. Qian, P. E. W. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhasaikhan, P. Nymadawa, and J. D. A. van Embden. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. *J. Clin. Microbiol.* **33**:3234–3238.
39. van Soolingen, D., P. de Haas, P. W. M. Hermans, P. Groenen, and J. D. A. van Embden. 1993. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *M. tuberculosis*. *J. Clin. Microbiol.* **31**:1987–1995.
40. van Soolingen, D., A. G. M. van der Zanden, P. E. W. de Haas, G. T. Noordhoek, A. Kiers, N. A. Foudraïne, F. Portaels, A. H. J. Kolk, K. Kremer, and J. D. A. van Embden. 1998. Diagnosis of *Mycobacterium microti* infections among humans by using novel genetic markers. *J. Clin. Microbiol.* **36**: 1840–1845.
41. van Soolingen, D., and P. W. M. Hermans. 1995. Epidemiology of tuberculosis by DNA fingerprinting. *Eur. Respir. J.* **8**:S649–S656.